Scientific Research Publishing

Open Journal of Statistics, 2026, 16(4), 284-298
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Extensions of the Mean Difference for the Lognormal Distribution

Fabio Manca[1], Angelo Vacca[2], Claudia Marin[1], Angelo Valerio[3], Elita Anna Sabella[2]

[1]Department of Education, Psychology, Communication Sciences, University of Bari Aldo Moro, Bari, Italy
[2]Department of Precision and Regenerative Medicine and Ionian Area, University of Bari Aldo Moro, Bari, Italy
[3]Research and Data Consultant, Bari, Italy
Email: fabio.manca@uniba.it, angelo.vacca@uniba.it, claudia.marin@uniba.it, angelovalerio.studio@gmail.com, elita.sabella@uniba.it

How to cite this paper: Manca, F., Vacca, A., Marin, C., Valerio, A. and Sabella, E.A. (2026) Extensions of the Mean Difference for the Lognormal Distribution. *Open Journal of Statistics*, **16**, 284-298.
https://doi.org/10.4236/ojs.2026.164013

Received: April 21, 2026
Accepted: August 2, 2026
Published: August 5, 2026

Abstract

This paper extends the closed-form formula of Gini's mean difference for the lognormal distribution, originally obtained by Girone and Manca (2016). The following are analyzed: 1) the asymptotic behavior of the scale parameter for $\gamma \to 0$ (degeneration of the distribution) and for $\gamma \to +\infty$ (unbounded dispersion); 2) the generalized formula with complete location parameter μ and scale parameter σ; 3) the mean difference conditioned on an interval; 4) the mean difference for the truncated lognormal. Applications in the field of medical sciences are also discussed, where the lognormal distribution is ubiquitous in the modeling of biomarkers and pharmacological concentrations. The results show that the original formula $\Delta = 2e^{\gamma^2/2}\mathrm{erf}\left(\gamma/2\right)$ for the standardized case admits natural extensions that significantly broaden its field of application, while preserving the analytical elegance of the basic formulation.

Keywords

Mean Difference, Lognormal Distribution, Truncated Distribution, Gini Index, Medical Sciences

1. Introduction

Gini's mean difference, introduced by Corrado Gini in 1912 [1], represents one of the most important measures of variability in statistical theory and applied practice. For a continuous random variable X with density function $f(x)$ and cumulative distribution function $F(x)$, defined on a support (a, b), the mean difference is defined as the expected value of the absolute difference between two independent observations X_1 and X_2 drawn from the same distribution: $\Delta = E[|X_1 - X_2|]$.

The mean difference possesses fundamental properties that make it particularly attractive as a measure of variability: it is translation-invariant (independent of the location parameter), it is homogeneous of degree one with respect to the scale parameter, and it is directly related to the Gini concentration index through the relation $G = \Delta/(2E[X])$, where $E[X]$ denotes the mean of the distribution. These properties, combined with its intuitive interpretability and robustness with respect to extreme values, make it an instrument of fundamental importance in fields ranging from economics to biometry, from epidemiology to reliability theory.

In addition to the variance, several alternative measures of variability and concentration have been proposed in the statistical and economic literature. Among them, Gini's mean difference occupies a central role, as it combines properties of dispersion and inequality measurement. Originally introduced by Corrado Gini, this measure has been shown to provide a more informative description of non-normal and skewed distributions compared to variance [2].

The Gini framework has been extensively developed in both theoretical and applied directions, particularly in the context of inequality measurement, where it is closely related to the Lorenz curve and concentration indices [3]. More recent contributions highlight the need for complementary measures capable of capturing different aspects of distributional heterogeneity [4].

Recent contributions have further advanced the study of the lognormal distribution and the mean difference in both theoretical and applied settings. Novi Inverardi and Tagliani [5] demonstrate that the lognormal distribution is uniquely characterized by its integer moments via maximum entropy. Dai and Shen [6] propose a robust two-quantile method for Gini coefficient estimation. Poudyal, Zhao, and Brazauskas [7] develop winsorized moment estimators for truncated and censored lognormal distributions. Kapera and Kobus [8] extend the Gini and mean log deviation indices to multivariate inequality of opportunity. Jokiel-Rokita and Piątek [9] construct estimators for Weibull parameters based on the nonparametric Gini coefficient. The present work contributes to this growing literature by providing analytical extensions of the lognormal mean difference formula.

In a series of previous studies, Girone and collaborators [10]-[12] obtained closed-form formulas for the mean difference for numerous continuous and discrete distributions, including the exponential, gamma, beta, Pareto, and geometric distributions. In particular, Girone and Manca [13] derived, through a procedure employing selected integrals of the error function, the remarkably simple formula for the mean difference of the standardized lognormal distribution (with location parameter $\theta = 0$):

$$\Delta = 2\mathrm{e}^{\gamma^2/2}\mathrm{erf}\left(\frac{\gamma}{2}\right) \tag{1}$$

where $\gamma > 0$ is the scale parameter (standard deviation of the logarithm of the variable)—a quantity that, in the generalized formulation of Section 4, will be

equivalently denoted by σ, both symbols representing the standard deviation of $\log X$ in their respective parametric contexts—and erf(·) denotes the error function, defined as:

$$\operatorname{erf}(x) = \frac{2}{\pi}\int_0^x e^{-t^2} dt. \tag{2}$$

This formula was obtained starting from the representation of the mean difference based exclusively on the cumulative distribution function, namely:

$$\Delta = 2\int_a^b F(x)\left[1 - F(x)\right]dx \tag{3}$$

and through a procedure involving six selected integrals of the error function, some of which are particularly complex and were computed by resorting to an integral given by Prudnikov, Brychkov and Marichev [14], which made it possible to reduce apparently intractable expressions to a formulation of remarkable simplicity and elegance[1].

The present work aims to extend this fundamental result in five complementary directions: 1) the rigorous analysis of the asymptotic behavior for $\gamma \to 0$ and $\gamma \to +\infty$, which provides a complete understanding of the mean difference under limiting conditions; 2) the generalization of the formula to the complete parameters μ (location) and σ (scale); 3) the formulation of the mean difference conditioned on an interval $[a, b]$; 4) the derivation of the mean difference for the truncated lognormal distribution; and 5) the discussion of applications in the field of medical sciences, where the lognormal distribution is omnipresent. Remark on notation. Throughout this paper, the following conventions are adopted to ensure notational clarity: 1) The symbol γ denotes the scale parameter (standard deviation of $\log X$) in the standardized formulation (Sections 1 - 3), while σ denotes the identical quantity in the general formulation (Sections 4 - 5); when $\mu = 0$, we have $\sigma \equiv \gamma$. 2) The symbol μ is reserved exclusively for the location parameter of the underlying normal distribution (the mean of $\log X$); the mean of the lognormal variable X is always denoted $E[X] = \exp(\mu + \sigma^2/2)$ to avoid ambiguity. 3) Context-specific subscripts are used for the mean difference where needed: Δ without subscript in Sections 1 - 3 refers to the standardized case (Formula (1)), $\Delta(\mu, \sigma)$ to the general case (Formula (10)), $\Delta_c[a, b]$ to the conditional mean difference on an interval (Formula (12)), and Δ_T to the truncated case (Formula (13)). The unqualified symbol Δ refers to the mean difference in its generic, context-dependent sense.

The remainder of this paper is organized as follows. Section 2 introduces the notation, definitions, and preliminary results used throughout the paper, including the formal definition of the lognormal distribution and the error function. Section 3 analyzes the asymptotic behavior of the mean difference for extreme

[1]In Formula (1), the parameter $\gamma > 0$ denotes the standard deviation of the logarithm of the variable, *i.e.*, $\gamma = \mathrm{sd}(\log X)$. This is the sole parameter of the standardized lognormal distribution (with location parameter $\mu = 0$). In the general formulation of Section 4, this parameter is denoted σ, with $\gamma \equiv \sigma$ when $\mu = 0$. The cumulative distribution function $F(x)$ and the support endpoints (a, b) appearing in Formula (3) refer to the general support of the distribution; for the standard lognormal, $a = 0$ and $b = +\infty$.

values of the scale parameter. Section 4 derives the generalized formula with complete parameters. Section 5 develops the conditional mean difference on an interval. Section 6 presents the mean difference for the truncated lognormal distribution. Section 7 discusses applications in the medical sciences. Finally, Section 8 presents conclusions, and Section 9 outlines limitations and directions for future research.

2. Notation and Definitions

This section collects the basic definitions, notation, and preliminary results that are used throughout the paper. The purpose is to establish a consistent notational framework and to make the paper self-contained.

***Definition 1* (*Lognormal Distribution*).**

A random variable X is said to follow a lognormal distribution with parameters μ and σ^2, written $X \sim \text{Lognormal}(\mu, \sigma^2)$, if $Y = \log X$ follows a normal distribution $N(\mu, \sigma^2)$. Equivalently, $X = \exp(Y)$ where $Y \sim N(\mu, \sigma^2)$. The probability density function of X is: $f(x;\mu,\sigma) = \left(1/\left(x\sigma\sqrt{2\pi}\right)\right)\exp\left(-(\log x-\mu)^2/\left(2\sigma^2\right)\right)$, for $x > 0$, and the cumulative distribution function is: $F(x;\mu,\sigma) = \Phi\left((\log x-\mu)/\sigma\right)$, where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution. The parameters μ and σ are, respectively, the mean and standard deviation of the underlying normal variable $Y = \log X$ (not of X itself). The mean and variance of X are: $E[X] = \exp(\mu + \sigma^2/2)$ and $\text{Var}(X) = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$.

***Remark 1* (*Standardized vs. General Lognormal*).**

When $\mu = 0$, the lognormal distribution depends only on the scale parameter σ. In the standardized case used in Sections 1 and 3, we write $\gamma = \sigma$ for the scale parameter (standard deviation of $\log X$). Setting $\mu = 0$ and $\sigma = \gamma$ in any general formula recovers the corresponding standardized result.

***Definition 2* (*Error Function*).**

The error function erf: $\mathbb{R} \to (-1, 1)$ is defined by: $\text{erf}(x) = \left(2/\sqrt{\pi}\right)\int_0^x \exp\left(-t^2\right)\text{d}t$. It is an odd function: $\text{erf}(-x) = -\text{erf}(x)$, with $\text{erf}(0) = 0$ and $\lim_{x\to\infty}\text{erf}(x) = 1$. The complementary error function is $\text{erfc}(x) = 1 - \text{erf}(x)$. The error function is related to the standard normal CDF by: $\Phi(x) = (1/2)\left[1+\text{erf}\left(x/\sqrt{2}\right)\right]$, so that $\text{erf}(x) = 2\Phi\left(x\sqrt{2}\right)-1$. This relationship is used extensively in the derivations that follow.

***Definition 3* (*Mean Difference*).**

For a continuous random variable X with cumulative distribution function $F(x)$ on a support (a, b), the mean difference (Gini's mean difference) is defined as: $\Delta = E[|X_1 - X_2|]$, where X_1 and X_2 are independent copies of X. An equivalent representation based solely on F is: $\Delta = 4\int_a^b F(x)\left[1-F(x)\right]\text{d}x$, as given by Formula (3) below. This integral form is the starting point for all derivations in this paper. The mean difference is translation-invariant, homogeneous of degree one, and provides a robust measure of variability that is less sensitive to extreme observations than the variance.

3. Asymptotic Behavior of the Mean Difference

3.1. Case $\gamma \to 0$: Degeneration of the Distribution

When the scale parameter γ tends to zero, the lognormal distribution degenerates toward a distribution concentrated at a single point. This behavior can be understood by analyzing the probabilistic structure of the distribution. If X follows a standardized lognormal distribution with scale parameter γ, then $Y = \log X$ has a normal distribution with mean zero and variance γ^2. Therefore, when $\gamma \to 0$, the variance of Y tends to zero, which implies, by Chebyshev's theorem, that Y converges in probability to zero.

Consequently, $X = e^Y$ converges in probability to $e^0 = 1$. Formally, for $\gamma \to 0$, the lognormal distribution converges weakly to the degenerate distribution δ_1 concentrated at the point $x = 1$ (that is, at the point $x = e^\theta$ in the general case with location parameter θ). This means that the entire probability mass progressively concentrates in an increasingly narrow neighborhood of the unit point, and the distribution tends to behave as a constant random variable.

Since for a degenerate distribution all observations assume the same value, the expected absolute difference between two independent observations must necessarily tend to zero. Analytically, this result is verified directly from Formula (1). For $\gamma \to 0$, since $\operatorname{erf}(0) = 0$ and $e^0 = 1$:

$$\lim_{\gamma \to 0} \Delta = 2 \cdot e^0 \cdot \operatorname{erf}(0) = 2 \cdot 1 \cdot 0 = 0. \tag{4}$$

More precisely, using the Taylor series expansion of the error function, $\operatorname{erf}(x) \approx 2x/\sqrt{\pi}$ for $x \to 0$, and the expansion of the exponential $e^{\gamma^2/2} \approx 1 + \gamma^2/2 + O(\gamma^4)$, the following asymptotic expansion is obtained:

$$\Delta \approx \frac{2}{\sqrt{\pi}} \gamma, \gamma \to 0 \tag{5}$$

This result confirms that the standardized mean difference $\Delta \to 0$ for $\gamma \to 0$, with a linear rate of convergence in the parameter γ. The coefficient $2/\sqrt{\pi} \approx 1.128$ quantifies the rate of decrease of the mean difference in the neighborhood of the origin. The result is fully consistent with the probabilistic interpretation: when the dispersion of the lognormal distribution vanishes, the absolute variability, as measured by the mean difference, also vanishes, and this occurs at a rate proportional to the dispersion parameter itself.

3.2. Case $\gamma \to +\infty$: Unbounded Dispersion

When the scale parameter γ tends to infinity, the lognormal distribution becomes increasingly dispersed, with a right tail that extends indefinitely and an increasing concentration of probability mass near zero. In this regime, the variance of $\log X$ diverges, producing an extremely positively skewed (right-skewed) distribution with a skewness coefficient that grows without bound.

To analyze the asymptotic behavior of the mean difference, we consider separately the two factors of Formula (1). The exponential factor and the error function factor behave as follows for $\gamma \to +\infty$:

$$e^{\gamma^2/2} \to +\infty, \operatorname{erf}\left(\frac{\gamma}{2}\right) \to 1 \tag{6}$$

The exponential factor $e^{\gamma^2/2}$ diverges at super-exponential speed, completely dominating the asymptotic behavior of the product, while the factor erf(γ/2) converges monotonically to 1 from below. Consequently, the product of the two factors diverges:

$$\Delta \to \infty. \tag{7}$$

This result is consistent with the fact that the mean of the standardized lognormal distribution is $E[X] = e^{\gamma^2/2}$, and therefore the mean difference, being a homogeneous function of the scale, must also diverge. More significant is the ratio of the mean difference to the mean of the distribution, which tends to the limiting value:

$$\frac{\Delta}{E[X]} = 2 \cdot \operatorname{erf}\left(\frac{\gamma}{2}\right) \to 2 \tag{8}$$

This indicates that, for large values of γ, the mean difference is approximately twice the mean, a result that reflects the extreme skewness of the lognormal distribution in the high-dispersion regime. In terms of the Gini concentration index, $G = \Delta/(2E[X])$, this implies that $G \to 1$ for $\gamma \to +\infty$, reaching the value of maximum concentration. This behavior is consistent with the well-known property that the Gini index of the lognormal distribution is given by $G = \operatorname{erf}(\sigma/2)$, which tends to 1 when $\sigma \to +\infty$.

4. Generalized Formula with Complete Parameters

Formula (1) was derived for the standardized lognormal distribution, with location parameter $\theta = 0$ and employing γ as the sole scale parameter. In statistical practice and applications, the complete lognormal distribution is characterized by two parameters: the location parameter μ (mean of the logarithm of the variable) and the scale parameter σ (standard deviation of the logarithm of the variable). A random variable X is said to have a lognormal distribution with parameters μ and σ^2 if $Y = \log X \sim N(\mu, \sigma^2)$. The density function of the general lognormal distribution is:

$$f(x) = \frac{e^{-\frac{(\log x - \mu)^2}{2\delta^2}}}{x\delta\sqrt{2\pi}}, x > 0. \tag{9}$$

The mean difference, being independent of the location parameter and homogeneous of degree one with respect to the scale parameter, can be generalized by exploiting these fundamental properties. Recall that if X is a random variable with mean difference $\Delta(X)$, and $Y = aX + b$ with $a > 0$, then $\Delta(Y) = a \cdot \Delta(X)$. This property is a direct consequence of the definition of the mean difference as the expected value of the absolute difference.

In the case of the lognormal distribution, if $X \sim$ Lognormal (μ, σ^2), we can write $X = e^{\mu+\sigma Z} = e^{\mu} \cdot e^{\sigma Z}$, where $Z \sim N(0, 1)$. The variable $W = e^{\sigma Z}$ follows a standardized

lognormal distribution whose scale parameter σ is precisely the same quantity as the parameter γ of Formula (1): both denote the standard deviation of $\log X$. Setting $\mu = 0$ and $\sigma = \gamma$ in the generalized formula recovers exactly Formula (1). Therefore, $X = e^{\mu} \cdot W$, and by the homogeneity property: $\Delta(X) = e^{\mu} \cdot \Delta(W) = e^{\mu} \cdot 2e^{\sigma^2/2}\text{erf}(\sigma/2)$.

This leads to the generalized formula for the lognormal mean difference:

$$\Delta(\mu,\sigma) = 2\exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \text{erf}\left(\frac{\sigma}{2}\right) \tag{10}$$

where: μ = location parameter (mean of the logarithm of the variable); σ = scale parameter (standard deviation of the logarithm of the variable); $\Delta(\mu, \sigma)$ = mean difference for the lognormal distribution with parameters μ and σ^2.

The term e^{μ} appears as a multiplicative factor arising from the exponential transformation of the location parameter. Note that Formula (10) reduces to (1) when $\mu = 0$ and $\sigma = \gamma$, confirming that γ and σ are notational variants of the same scale parameter and making the relationship between the two formulations fully transparent. Furthermore, from Formula (10) and the well-known expression for the mean of the lognormal distribution, $E[X] = e^{\mu+\sigma^2/2}$, one immediately obtains the expression for the Gini concentration index:

$$G = \frac{\Delta}{2E[X]} = \text{erf}\left(\frac{\sigma}{2}\right) \tag{11}$$

which depends solely on the scale parameter σ and not on the location parameter μ, confirming a well-known property of the lognormal distribution in concentration theory.

5. Mean Difference Conditioned on an Interval

In many statistical applications, it is of interest to measure the variability not of the entire distribution, but of a subpopulation defined by constraints on the observable values. This need arises frequently in economic contexts (analysis of income variability within specific brackets), in biomedical contexts (variability of a biomarker within a clinically relevant interval), and in industrial contexts (variability of a quantity within tolerance specifications).

The mean difference conditioned on an interval $[a, b]$ represents the natural extension of the concept of the mean difference to the case where only observations falling between the limits a and b are considered. Let X be a random variable with cumulative distribution function $F(x)$ and let $0 < F(a) < F(b) < 1$. The conditional distribution of X given $a \le X \le b$ has density $f(x|a \le X \le b) = f(x)/[F(b) - F(a)]$ for $x \in [a, b]$ and conditional cumulative distribution function $F_c(x) = [F(x) - F(a)]/[F(b) - F(a)]$.

[2]Let $X \sim$ Lognormal(μ, σ^2). Then the mean difference of X is given by: $\Delta(\mu, \sigma) = 2\exp(\mu + \sigma^2/2)\cdot\text{erf}(\sigma/2)$, as stated in Formula (10) below. Since $X = \exp(\mu + \sigma Z) = \exp(\mu)\cdot\exp(\sigma Z)$ where $Z \sim N(0,1)$, and $W = \exp(\sigma Z)$ follows a standardized lognormal distribution, the homogeneity property $\Delta(X) = \exp(\mu)\cdot\Delta(W)$ together with Formula (1) yields the result.

Applying the formula for the mean difference based on the cumulative distribution function to the conditional distribution, and taking into account that $F_c(x)\left[1-F_c(x)\right]=\left[F(x)-F(a)\right]\left[F(b)-F(x)\right]\Big/\left[F(b)-F(a)\right]^2$, the following general formula for the conditional mean difference, hereafter denoted $\Delta_c[a, b]$, is obtained:

$$\Delta_{[a,b]}=\frac{4}{\left[F(b)-F(a)\right]}\int_a^b\left[F(x)-F(a)\right]\left[F(b)-F(x)\right]\mathrm{d}x \tag{12}$$

where $F(a)$ and $F(b)$ are the values of the cumulative distribution function at the endpoints of the interval. The factor $4/[F(b) - F(a)]^2$ serves as a normalization that accounts for the probability mass contained in the interval[3].

Formula (12) has a particularly elegant structure: the integrand $[F(x) - F(a)]$ $[F(b) - F(x)]$ is a non-negative function that vanishes at the endpoints of the interval and reaches its maximum at the conditional median. This structure reflects the fact that the pairs of observations with the largest expected difference are those located on opposite sides of the median of the conditional distribution.

In the context of the lognormal distribution, this formulation is particularly relevant when analyzing subgroups defined by economic thresholds (income brackets), biological thresholds (concentration intervals), clinical thresholds (diagnostic value ranges), or environmental thresholds (exposure brackets). The formula allows quantifying intra-group variability analytically, avoiding the approximations that would be necessary with numerical methods.

6. Mean Difference of the Truncated Lognormal Distribution

Truncation is frequent in practical applications where extreme values are physically impossible, unobservable, or subject to censoring. The lognormal distribution truncated on the interval $[a, b]$ finds application in numerous contexts: biochemical concentrations (which cannot be negative and have physiological upper bounds), incomes (bounded below and above in sample surveys), survival times (truncated by the observation period of the study), and particle sizes (limited by instrumental detection thresholds).

The formula for the mean difference of the truncated lognormal is obtained by specializing the conditional mean difference Formula (12) to the lognormal distribution. Through the substitution of the lognormal cumulative distribution function $F(x) = \Phi((\log x - \mu)/\sigma)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and the change of variable $y = (\log x - \mu)/\sigma$, the limits of integration are transformed into standardized limits. After appropriate algebraic simplifications, the following formula for the truncated mean difference Δ_T is obtained:

[3]Let X be a continuous random variable with CDF $F(x)$ on support (a_0, b_0). The mean difference of X conditioned on the interval $[a, b] \subseteq (a_0, b_0)$ is:

$\Delta_c[a,b]=\left(4\Big/\left[F(b)-F(a)\right]^2\right)\int_a^b\left[F(x)-F(a)\right]\left[F(b)-F(x)\right]\mathrm{d}x$, as given in Formula (12) below. The result follows from applying the integral representation (3) to the conditional distribution of X given $a \le X \le b$, whose CDF is $F_{[a,b]}(x) = [F(x) - F(a)]/[F(b) - F(a)]$, and simplifying.

$$\Delta^{(T)} = \frac{2e^{\mu+\sigma^2/2}}{\Phi(\beta^*)-\Phi(\alpha^*)}\left[\operatorname{erf}\left(\frac{\beta^*-\sigma}{\sqrt{2}}\right)-\operatorname{erf}\left(\frac{\alpha^*-\sigma}{\sqrt{2}}\right)\right]-\Delta_0 \tag{13}$$

where the standardized limits are defined as:

$$\alpha^* = (\log(a-\mu))/\sigma,\ \beta^* = (\log(b-\mu))/\sigma \tag{14}$$

and the other variables are: $[a, b]$ = truncation interval; $\Phi(\cdot)$ = cumulative distribution function of the standard normal distribution; Δ_0 = corrective term for the truncation, defined as the residual adjustment that ensures the truncated mean difference Δ_T converges to the unrestricted $\Delta(\mu, \sigma)$ when the truncation interval expands to $(0, +\infty)$[4].

The derivation proceeds from the conditional mean difference formulation (Section 5), specialized to the lognormal distribution with parameters μ and σ. The substitution of the lognormal cumulative distribution function into the integral of Formula (12) and the application of the change of variable lead to integrals involving the error function, analogous to those used by Girone and Manca [13] in the derivation of the original formula.

The corrective term Δ_0 represents the boundary correction that accounts for the effects introduced by the truncation on the mean difference. Specifically, Δ_0 is defined as the double integral

$$\Delta_0 = \left(4/\left[\Phi(\beta^*)-\Phi(\alpha^*)\right]^2\right)\iint_{[a,b]^2}\left|\Phi((\log x-\mu)/\sigma)-\Phi((\log y-\mu)/\sigma)\right| \cdot\left[f(x)f(y)-f_T(x)f_T(y)\right]\mathrm{d}x\mathrm{d}y,$$

where f_T denotes the truncated density. In general, Δ_0 does not admit a closed-form expression and must be evaluated numerically; however, it vanishes as the truncation interval expands to the full support, since $f_T \to f$. It ensures that Δ_T converges correctly to the non-truncated mean difference $\Delta(\mu, \sigma)$ of Formula (10) when the truncation limits tend respectively to 0 and $+\infty$, that is, when $\alpha^* \to -\infty$ and $\beta^* \to +\infty$. In that limit, indeed, $\Phi(\beta^*) - \Phi(\alpha^*) \to 1$ and the corrective term vanishes, recovering exactly Formula (10).

Formula (13) thus extends the fundamental Girone-Manca result to the case of the truncated lognormal distribution, maintaining the analytical structure based on the error function that characterizes the original formulation.

7. Applications in Medical Sciences

The extensions of the lognormal mean difference presented in the preceding sections find numerous and significant applications in the field of medical sciences, where the lognormal distribution represents one of the most frequently employed probabilistic models for describing the variability of biological phenomena [15].

[4]Let $X \sim$ Lognormal(μ, σ^2) truncated to the interval $[a, b]$ with $0 < a < b$. Then the mean difference of the truncated distribution is given by Formula (13) below, expressed in terms of the standardized truncation limits $\alpha = (\log a - \mu)/\sigma$ and $\beta = (\log b - \mu)/\sigma$, the standard normal CDF $\Phi(\cdot)$, and a boundary correction term Δ_0. The result is obtained by specializing Proposition 2 to the lognormal CDF $F(x; \mu, \sigma) = \Phi((\log x - \mu)/\sigma)$ and evaluating the resulting integral using the substitution $u = (\log x - \mu)/\sigma$, together with known integrals of the error function [9].

7.1. Biomarkers and Biochemical Concentrations

Many biological variables follow a lognormal distribution: serum concentrations of proteins (such as C-reactive protein, CRP), hepatic enzymes (ALT, AST), hormones (cortisol, testosterone, TSH), antibodies (IgG, IgM, IgA), blood glucose levels, plasma drug concentrations, and numerous other biomarkers [15]. The generalized mean difference (10) allows quantifying inter-individual variability in these measurements, providing a robust measure of dispersion that, unlike the variance, is not excessively influenced by the extreme values typically present in biochemical distributions.

In particular, in laboratory practice, reference intervals for biomarkers are typically defined as percentiles of the distribution in the healthy population. The conditional mean difference (12), calculated over the reference interval, provides a measure of "physiological" variability that excludes pathological values and allows more appropriate comparisons between different populations.

Among the biomarkers that most frequently exhibit a lognormal distribution, prostate-specific antigen (PSA) represents an emblematic case. In the healthy male population, serum PSA concentrations show a marked positive skewness, with a right tail extending toward elevated values associated with pathological conditions such as benign prostatic hyperplasia and prostate carcinoma. The conventional diagnostic threshold of 4.0 ng/mL, used as a cut-off value for oncological screening, is located in the upper portion of the lognormal distribution of normal values, and the mean difference Δ calculated on the distribution of sub-threshold values provides a measure of physiological variability that can be employed to refine diagnostic criteria as a function of the patient's age and ethnicity.

Similarly, cardiac troponin (cTnI and cTnT), used as a primary biomarker for the diagnosis of acute myocardial infarction, exhibits a lognormal distribution in the general population, with extremely low values in the majority of healthy subjects and increases of several orders of magnitude in the presence of myocardial damage. Serum creatinine levels, a fundamental indicator of renal function, follow an analogous distributional pattern, with the mean difference being particularly informative in quantifying inter-individual variability while accounting for the multiplicative nature of glomerular filtration processes. The hepatic enzymes alanine aminotransferase (ALT) and aspartate aminotransferase (AST) also follow a lognormal distribution, as do the tumor markers CA-125, employed in the monitoring of ovarian carcinoma, and alpha-fetoprotein (AFP), used in the surveillance of hepatocellular carcinoma.

The fundamental reason why these biomarkers follow a lognormal distribution lies in the nature of the underlying biological processes. Biochemical concentrations in plasma are the result of cascades of enzymatic reactions, processes of synthesis, secretion, distribution, metabolism, and elimination that operate in a multiplicative rather than additive manner. According to the multiplicative central limit theorem, the product of a large number of independent positive random variables converges to a lognormal distribution, regardless of the distribution of

the individual component variables [15]. This principle explains the ubiquity of the lognormal distribution in biomedical sciences: each metabolic step modifies the concentration by a multiplicative factor, and the overall effect is a log-normal distribution of the values observed in the population.

From a clinical standpoint, the use of the mean difference Δ in place of the standard deviation offers specific advantages in characterizing biomarker variability. The standard deviation, calculated on the original scale of concentrations, is heavily influenced by extreme values and provides distorted information when the distribution is markedly skewed. The mean difference, being based on the absolute differences between all pairs of observations, is inherently more robust and captures the effective dispersion of the data in a manner more representative of clinically relevant variability. Furthermore, the ratio $\Delta/E[X]$, where $E[X]$ denotes the mean of the distribution, provides an alternative coefficient of variability to the classical coefficient of variation $\mathrm{CV} = \mathrm{SD}(X)/E[X]$, with superior robustness properties for heavy-tailed distributions, such as the lognormal.

The definition of diagnostic thresholds and reference intervals benefits significantly from the analytical formulation proposed here. Reference intervals, traditionally constructed as the central 95th percentile of the distribution of values in the healthy population, can be supplemented with the conditional mean difference (12) calculated over the interval itself, yielding a more complete characterization of intra-interval variability. This approach is particularly useful in comparing different populations (by sex, age, ethnicity, or physiological conditions such as pregnancy), where not only the limits of the reference interval, but also the structure of the internal variability may differ in a clinically significant manner.

To illustrate the practical relevance of the proposed formulation, consider a biomarker whose distribution in a reference population can be reasonably approximated by a lognormal distribution with parameters $\mu = 0$ and $\sigma = 0.5$.

Using the generalized formula for the mean difference, we obtain:

$$\Delta(\mu,\sigma) = 2\exp\left(\mu + \frac{\sigma^2}{2}\right)\cdot \operatorname{erf}\left(\frac{\sigma}{2}\right) \tag{15}$$

which yields $\Delta \approx 0.84$. Since the mean of the distribution is $E[X] = \mathrm{e}^{\mu+\sigma^2/2} \approx 1.13$, the ratio $\Delta/E[X] \approx 0.74$ provides a normalized measure of variability.

For comparison, consider a more dispersed scenario with $\sigma = 1$. In this case, the mean increases to $E[X] \approx 1.65$, while the mean difference rises to $\Delta \approx 2.05$, leading to $\Delta/E[X] \approx 1.24$.

These results highlight how the mean difference captures the rapid increase in variability associated with higher dispersion in lognormal models, reflecting the growing asymmetry and heavy-tailed behavior typical of biological measurements.

7.2. Pharmacological Dosages and Pharmacokinetics

The pharmacokinetics of many drugs is described by lognormal models for

plasma concentrations [16]. Inter-individual variability in pharmacokinetics is a crucial factor in the design of dosing regimens and in the evaluation of drug safety.

The conditional mean difference (12) finds specific application in the evaluation of the variability of concentrations within the therapeutic window, that is, the concentration interval between the minimum effective concentration (MEC) and the minimum toxic concentration (MTC). The analytical quantification of this intra-therapeutic variability is of fundamental importance for the optimization of dosing regimens and for the personalization of pharmacological therapy.

The fundamental pharmacokinetic parameters—the maximum plasma concentration (Cmax), the area under the concentration-time curve (AUC), systemic clearance (CL), the volume of distribution (Vd), and the elimination half-life ($t_{1/2}$)—typically exhibit a lognormal distribution in the population [16]. This distributional property is a direct consequence of the nature of pharmacokinetic processes: gastrointestinal absorption, hepatic first-pass metabolism, plasma protein binding, and glomerular filtration operate as multiplicative factors on the drug concentration, generating inter-individual variability that is expressed on a logarithmic scale. In particular, the hepatic first-pass effect, mediated by the cytochrome P450 enzyme system, introduces substantial variability due to the genetic polymorphism of metabolizing enzymes (CYP2D6, CYP3A4, CYP2C19), with differences between poor, intermediate, extensive, and ultra-rapid metabolizers that translate into variations in Cmax and AUC by a factor ranging from 2 to over 10.

The generalized mean difference (10) finds direct application in the quantification of inter-individual variability of these pharmacokinetic parameters. For a drug with log-normally distributed bioavailability parameters, the formula $\Delta = 2e^{\mu+\sigma^2/2} \cdot \operatorname{erf}\left(\sigma/2\right)$ allows the analytical calculation of the expected mean difference between two randomly selected subjects from the population, providing complementary and more intuitive information compared to the coefficient of variation traditionally reported in pharmacokinetic studies. This is particularly relevant for drugs with a narrow therapeutic index (warfarin, digoxin, phenytoin, lithium, cyclosporine, tacrolimus), where small concentration variations can determine the transition from therapeutic efficacy to toxicity or inefficacy.

In the context of Therapeutic Drug Monitoring (TDM), the mean difference conditioned on the therapeutic window [MEC, MTC] assumes a central role. TDM involves the periodic measurement of plasma drug concentrations with the objective of maintaining levels within the therapeutic range. Formula (12), applied to the interval [MEC, MTC], allows quantifying the expected intra-therapeutic variability in the population, information that is essential for the design of monitoring protocols and for defining the optimal frequency of blood sampling. Furthermore, the comparison between the mean difference over the entire distribution and the mean difference conditioned on the therapeutic window provides a quantitative measure of the proportion of clinically relevant variability, distinguishing it from the total variability that includes sub-therapeutic and supra-therapeutic values.

Bioequivalence studies, which are fundamental for the regulatory approval of generic drugs, are explicitly based on the assumption of lognormality of pharmacokinetic parameters. The guidelines of the FDA (Food and Drug Administration) and the EMA (European Medicines Agency) require that the ratio of pharmacokinetic parameters (Cmax and AUC) of the generic drug to the reference drug be analyzed on a logarithmic scale, with 90% confidence intervals that must fall within the 80% - 125% range. In this context, the mean difference Δ on the logarithmic scale of the parameters offers a complementary indicator of the variability of the bioequivalence ratio, while the truncated lognormal Formula (13) can be employed to characterize the variability when data are subject to analytical quantification limits or when pharmacokinetic outliers are excluded from the dataset.

Population pharmacokinetics (PopPK), which employs nonlinear mixed-effects models to describe inter- and intra-individual variability in the pharmacokinetic response, provides estimates of the location parameter μ and the scale parameter σ of the lognormal distribution of pharmacokinetic parameters in the population. These estimated parameters can be directly inserted into Formula (10) to obtain the expected mean difference in the patient population studied. The PopPK approach also identifies covariates (body weight, renal and hepatic function, age, sex, metabolic genotype) that explain part of the inter-individual variability, and the mean difference can be calculated conditionally on the subgroups defined by these covariates, allowing a more precise quantification of residual variability and supporting personalized dosing strategies based on the individual patient's characteristics.

Consider a pharmacokinetic variable (e.g., plasma concentration) following a lognormal distribution with parameters $\mu = 1$ and $\sigma = 0.6$. Suppose that the therapeutic window is defined by the interval $[a, b] = [1, 5]$.

The conditional mean difference over this interval can be evaluated using the proposed formulation based on the cumulative distribution function. Numerical evaluation shows that the conditional mean difference is substantially lower than the unconditional one, reflecting the restriction to clinically relevant values.

This reduction quantifies the extent to which variability outside the therapeutic window contributes to overall dispersion. From a clinical perspective, this provides a meaningful measure of intra-therapeutic variability, which may be useful in the design of dosing strategies and monitoring protocols.

More generally, the comparison between conditional and unconditional mean differences offers a quantitative tool to distinguish clinically relevant variability from extreme or non-informative observations.

8. Conclusions

In this work, four significant extensions of the mean difference formula for the lognormal distribution obtained by Girone and Manca [13] have been presented. The analysis of the asymptotic behavior has shown that the mean difference tends to zero for $\gamma \to 0$ (consistently with the degeneration of the distribution toward a

degenerate distribution) with a linear convergence rate proportional to $2\sqrt{2/\pi}\gamma$, and diverges to $+\infty$ for $\gamma \to +\infty$ (consistently with the divergence of the scale of the distribution) with a ratio $\Delta/E[X]$ that tends to 2 (where Δ here denotes the standardized mean difference).

The generalized formula with complete parameters μ and σ, obtained through the homogeneity property of the mean difference, extends the original result to the non-standardized lognormal model, making the formula directly applicable to empirical data parameterized in the usual form. The expression for the Gini index $G = \text{erf}(\sigma/2)$ that follows confirms well-known results in the literature on concentration.

The conditional mean difference and the formula for the truncated lognormal broaden the field of application to practical contexts where data are naturally limited or segmented. In particular, the formula for the truncated distribution, which extends the Girone-Manca result while maintaining the analytical structure based on the error function, represents a useful tool for the analysis of censored or truncated data.

The applications in medical sciences illustrate the practical relevance of these analytical results in fields ranging from clinical biochemistry to pharmacokinetics. The formulas derived here preserve the elegance and simplicity of the original formulation, confirming the centrality of the error function in the characterization of the variability of the lognormal distribution.

9. Limitations and Future Research

Despite the analytical generality of the results, some limitations of the present study should be acknowledged. The proposed formulations are derived under the assumption of exact lognormality, which, although widely supported in many applied contexts, may not fully capture deviations observed in empirical data. In addition, the results are obtained within a univariate framework and do not directly extend to multivariate settings, where dependence structures may play a relevant role.

Future research may address these limitations by extending the analysis to multivariate lognormal models and to alternative heavy-tailed distributions. Further developments could also include the empirical validation of the proposed measures on real datasets, as well as the comparison with other dispersion and inequality indicators in applied contexts. Such extensions would contribute to a deeper understanding of variability in complex stochastic systems.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Gini, C. (1912) Variabilità e Mutabilità. Contributo allo studio delle distribuzioni e delle relazioni statistiche. Tipografia di Paolo Cuppini.

[2] Yitzhaki, S. (2003) Gini's Mean Difference: A Superior Measure of Variability for Non-Normal Distributions. *METRON—International Journal of Statistics*, **61**, 285-316.

[3] Giorgi, G.M. ang Gigliarano, C. (2017) The Gini Concentration Index: A Review of the Inference Literature. *Journal of Economic Surveys*, **31**, 1130-1148. https://doi.org/10.1111/joes.12185

[4] Ebert, U. (2010) The Decomposition of Inequality Reconsidered: Weakly Decomposable Measures. *Mathematical Social Sciences*, **60**, 94-103. https://doi.org/10.1016/j.mathsocsci.2010.05.001

[5] Novi Inverardi, P.L. and Tagliani, A. (2024) The Lognormal Distribution Is Characterized by Its Integer Moments. *Mathematics*, **12**, Article 3830. https://doi.org/10.3390/math12233830

[6] Dai, P. and Shen, S. (2025) Estimation of the Gini Coefficient Based on Two Quantiles. *PLOS ONE*, **20**, e0318833. https://doi.org/10.1371/journal.pone.0318833

[7] Poudyal, C., Zhao, Q. and Brazauskas, V. (2023) Method of Winsorized Moments for Robust Fitting of Truncated and Censored Lognormal Distributions. *North American Actuarial Journal*, **28**, 236-260. https://doi.org/10.1080/10920277.2023.2183869

[8] Kapera, M. and Kobus, M. (2024) The Gini and Mean Log Deviation Indices of Multivariate Inequality of Opportunity. *Econometrics*, **12**, Article 10. https://doi.org/10.3390/econometrics12020010

[9] Jokiel-Rokita, A. and Piątek, S. (2022) Estimation of Parameters and Quantiles of the Weibull Distribution. *Statistical Papers*, **65**, 1-18. https://doi.org/10.1007/s00362-022-01379-9

[10] Girone, G. and D'Uggento, A.M. (2016) About the Mean Difference of the Inverse Normal Distribution. *Applied Mathematics*, **7**, 1504-1509. https://doi.org/10.4236/am.2016.714130

[11] Girone, G., Massari, A. and Mazzitelli, D. (2015) More on the Mean Difference of Continuous Distributive Models. *Proceedings SIS Conference* "*Statistics and Demography*: *The Legacy of Corrado Gini*", Treviso, 9-11 September 2015. https://www.scirp.org/reference/referencespapers?referenceid=1765353

[12] Girone, G., Manca, F. and D'Uggento, A.M. (2015) The Mean Difference of Discrete Distribution Models. *Proceedings SIS Conference* "*Statistics and Demography*: *The Legacy of Corrado Gini*", Treviso, 9-11 September 2015, 1-6.

[13] Girone, G. and Manca, F. (2016) The Mean Difference for Lognormal Distribution. *Applied Mathematics*, **7**, 824-828. https://doi.org/10.4236/am.2016.79073

[14] Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. (1986) Integrals and Series, Vol. 2: Special Functions. Gordon and Breach Science Publishers.

[15] Limpert, E., Stahel, W.A. and Abbt, M. (2001) Log-Normal Distributions across the Sciences: Keys and Clues. *BioScience*, **51**, 341-352. https://doi.org/10.1641/0006-3568(2001)051[0341:lndats]2.0.co;2

[16] Lacey, L.F., Keene, O.N., Pritchard, J.F. and Bye, A. (1997) Common Noncompartmental Pharmacokinetic Variables: Are They Normally or Log-Normally Distributed? *Journal of Biopharmaceutical Statistics*, **7**, 171-178. https://doi.org/10.1080/10543409708835177